一种基于全天相机云图的云量测量指标*

张雨昕 1，邱波 1，石超君 1，李梦慈 1，相冠杰 1

（1.河北工业大学，天津 300401）

摘要：全天相机云图是天文界监测云量的重要手段，但没有确定的云量计算指标。因此本文提出了一种新的云量测量量化指标——云分布密度，并基于该指标建立全天相机云图自动分类系统。首先，对云图进行去噪，利用 Otsu 算法分割云区域；然后对去除背景的云区域图像使用本文提出的云分布密度计算云量；最后使用四种传统的分类器（支持向量机，K 最近邻，决策树和随机森林）根据计算数值进行自动分类并评估各分类器的性能。结果表明，本文提出的云分布密度可作为评判全天相机云图云量的数值指标；基于云分布密度建立的云图自动分类系统实现了较高的识别准确率，其中随机森林法得到了最好的分类效果——各类云图的识别准确率均达到 95%以上。

关键字：云分布密度，全天相机云图，TMT，云量测量

中图分类号： P412. 15

0 引言

全天相机[1]拍摄的云图是目前天文台判别地球大气条件的重要依据，是天文观测工作不可缺少的重要工具，特别是在天文台的选址工作中至关重要，而云量是全天相机云图反映的云的重要信息之一。云量指的是特定类型的云（部分云量）或全部类型的云（总云量）覆盖天空的估计比例，目前对于云量计算的研究大部分是基于云检测来实现的，一类是根据云点与总像元点之间的比值来计算云量。比如国际卫星测云气候学计划（ISCCP）根据像元辐射值与晴空辐射之间的关系，判断像元是云点还是非云点[2]。阈值法根据天空与云的红蓝波段比值的不同，利用固定阈值[3]、全局自适应阈值[4]、局部阈值[5]对云点进行判断。超高分辨率辐射仪第一阶段云量算法(CLAVR-1)错误!未找到引用源。将像元分为无云、混合和有云三类；中等分辨率成像光谱仪云量算法（MODIS）[2]将像元分为确定云、可能云、可能晴空、晴空四类。此外还有超高分辨率辐射仪拓展云量算法（CLAVR-X）[6]，高分辨率红外测深仪云量算法（HIRS）[7]方法等。但是它们都是将像元分为有限的几个类别，做一个定性的分析，这无疑是很粗糙的。此外还有基于特征的云检测分类算法的研究如随机森林、支持向量机、贝叶斯分类器[8]、K-means++[9]、神经网络[10]等，主要基于颜色[11-14]或纹理特征[12]进行云区域分割。近年来深度学习也在云检测领域得到了广泛应用，比如一种基于深度残差全卷积网络[14]的云检测方法实现了对遥感影像云层目标像素级别的分割，像素精度达到了 93.33%；还有一种基于带权重的多尺度融合分割网络的遥感云图云检测方法 0，其像素精度可达 95.39%；

---

*：基金项目：国家自然科学天文联合基金(U1931134)资助.
作者简介：张雨昕，女，硕士，研究方向：机器视觉与模式识别。
通讯作者：邱波，男，教授，研究方向：模式识别。

某改进的全卷积神经网络实现了日间地基云图和夜间地基云图的云区域分割[15]，像素精度达到了 94.59%。另一类就是根据像元辐射量与反射率的比值来计算等效云量，但是其仅仅是对云覆盖面积的一个反应。

因此无论是云量还是等效云量，现阶段云量的研究仅仅关注于云层的覆盖面积而忽略了云的分布密集程度。

目前天文台的云量观测是使用 30 米口径望远镜（Thirty Meters Telescope, TMT）中用到的人工判读云图法：分别在天顶距 44.7 °和 65 °处画圆（在图 1 中用蓝色和绿色标注），将全天云量分为 Clear（外圈以内无云）、Covered（外圈以内厚云覆盖超过 50%）、Inner（内圈有云）、Outer（内圈无云，外圈到内圈之间有云）四类。

图 1 全天相机云图区域划分

Fig. 1. Division of ASI image region.

*：基金项目：国家自然科学天文联合基金(U1931134)资助.
作者简介：张雨昕，女，硕士，研究方向：机器视觉与模式识别。
通讯作者：邱波，男，教授，研究方向：模式识别。

本质上来说，TMT 方法主要是基于 ASI 图像中云区域的比例来测量云量，对云分布的密集程度考虑不够充分，并且没有直观可靠的定量指标。

针对上述问题，本文提出了一种新的云量测量指标——云分布密度（Cloud Distribution Density of ASI Images，ASICDD），用于描述围绕云图中心点的全天云量的分布情况。本文的主要创新点有：（1）将云点与云图中心点之间的距离变量引入云量计算中，定义了云分布密度的计算模型；（2）重新定义云面积的计算模型，引入云层厚度变量；（3）提出了一种基于云分布密度和云面积的云图自动分类系统，使用四种传统分类算法（决策树，KNN，SVM 和随机森林）并选取 1511 张云图数据进行实验分析。结果表明，随机森林（RF）的分类结果是最好的，并且在各种类型中具有很高的识别精度。

本文使用的云图数据来源于中国科学院国家天文台光学天文学重点实验室，大小为 1411×1411 像素。数据集中包含 1511 张全天相机云图，包括 295 张 Clear，696 张 Covered，277 张 Inner 和 243 张 Outer。

1 云分布密度的定义

1.1 云分布密度计算模型

云分布密度的提出主要是为了说明云区域距云图中心的远近对天文观测结果的不同影响，距离中心越近，影响越大，反之则越小。云分布密度就是对这种影响进行量化表示。云分布密度有三个决定要素：面积、厚度和距云图中心点的距离。在归一化处理后的云区域灰度图的基础上，将云区域划分为 N 个单位云区域 x_i（i=1,2,...,N），x_i 的面积为 s_i，平均云层厚度为 w_i，云区域中心点距云图中心点的距离为 d_i，一张 ASI 的云分布密度公式可如下表示：

$$\text{ASICDD} = \sum_{i=1}^{N} s_i w_i f(d_i) \qquad (1)$$

其中 f 是以 d_i 为自变量的云区域的分布密度函数。由于设备和裁剪的问题，距离参数不是固定的，因此取相对距离 $d_i^* = d_i/R$，其中 R 表示 ASI 有效区域半径。

云分布密度主要探究的是寻找合适的函数来表达云区域距云图中心点距离大小的影响。根据云分布密度的定义，相同面积大小的云区域距离中心点越近对观测影响越大，则定义式（1）中的距离分布函数 f 是随 d_i 递减的函数，在云图中心点分布密度最大。由于本文对全天相机云图的分类是基于 TMT 的判断标准，则在参数设定上考虑内圈和外圈的影响。经过多次试验和与人工观测的结果对比研究，人为设定云图中心点处分布密度为 1，内圈处为 0.7，外圈处为 0，经 MATLAB 曲线拟合工具箱线性拟合后 f 如图 2 所示，公式表达如式（2），其中正弦项是为了修正拟合的二次函数，使其满足递减条件。

$$f(d_i^*) = -24.33\sin(d_i^* - \pi) + 1.228(d_i^* - 10)^2 - 121.8 \qquad (2)$$

在实际运算中，本文以每个像素点为计算单位，可使 ASICDD 测量精确到像素级别，则式（1）中 x_i 是云图像素点，面积 s_i 为 1，由于云图中天空呈蓝色而云呈白色，且云层越厚云越白，即云层厚度与云点的灰度值呈正比，因此云层厚度 w_i 由灰度值 V_i 表示

$$w_i = V_i/255 \qquad (3)$$

其中 V_i 的取值范围为 0~255。综合式（1）（2）（3）可得到基于像素的 ASICDD 计算模型

$$\text{ASICDD} = \sum_{i=1}^{N_{cloud}} \frac{V_i f(d_i/R)}{255} \qquad (4)$$

其中，V_i 表示云图像素值，d_i 表示该像素点距云图中心点的距离，N_{cloud} 表示云点总数。计算结果进行归一化处理，使云分布密度值映射到 0-100 之间。图 3 表示 Clear、Covered、Inner、Outer 四类云图的云分布密度的分布情况，可以看出各类云图有较明显的分界，说明云分布密度可作为基于云量的云图分类的标准。

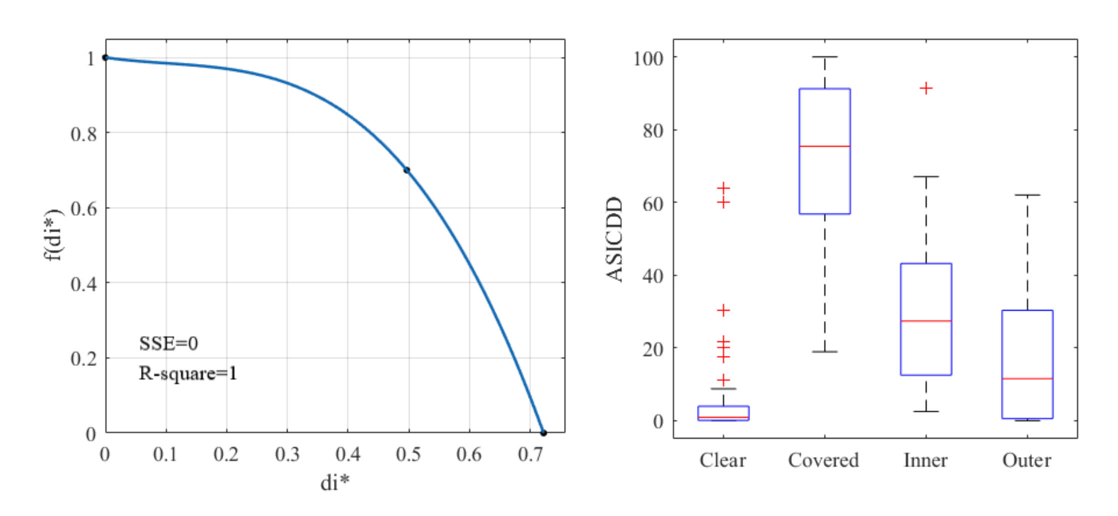


图 2 距离分布函数 *f* 与相对距离 di*的关系

Fig.2 The relationship between distance distribution function *f* and relative distance di*.

图 3 四类云图的云分布密度分布

Fig.3 ASICDD of four categories of ASI images.

1.2 云面积计算模型

目前国际上主要是针对卫星云图根据云检测得到的云像元，将云像元分为不同的种类，统计计算每种像元的概率，从而得到总云量，以此反映云层的覆盖面积。计算的方法主要有 ISCCP、CLAVR-1、CLAVR-X、MODIS 等。本文在 CLAVR-X 的基础上，以云图每个像素点定义云面积的计算模型，所有云像素点的加权面积和与整个云图有效区域的面积比，作为代表云面积（Scloud）的指标。以云层厚度作为云点面积权重，即根据灰度值定义权重参数，灰度值 V_i 的取值范围为 0~255，将其线性划分为 255 个等级。云面积计算公式如式（5）所示，其中 S_{global} 是云图有效区域面积，以像素为单位，本文中是 S_{global}=1411×1411 像素。

$$\text{Scloud} = \frac{\sum_{i=1}^{N_{cloud}} V_i}{255 S_{global}} \tag{5}$$

2 基于云量计算的云图分类系统

国内已有关于云图自动分类的相关研究：文献[17]结合卫星云图的纹理特征和光谱特征采用 SVM 算法实现卫星云图的自动分类；文献[18]提出一种针对全天相机云图的云量自动计算方法，并依据 TMT 判读全天相机云图的方法对云图进行自动分类。但这两种方法没有对全天相机云图的云量特征进行深入描述。本文以云分布密度和云面积来描述云图的云量特征，并结合机器学习算法构建基于云量计算的云图分类系统，图 4 是自动分类系统的流程框图。云图经去噪、分割出云区域图像后，在此基础上进行云量的数值计算，最后以云分布密度和云面积为特征的分类器对云图进行自动分类。

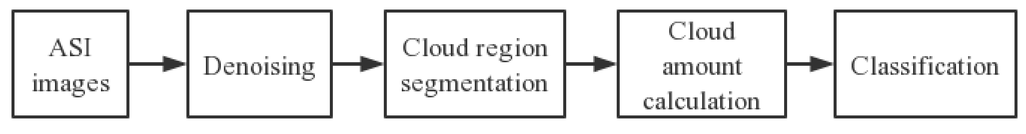


图 4 云图自动分类系统流程图

Fig.4 Cloud map automatic classification system flow chart

2.1 云图去噪

云图的固定噪声[19]会影响云检测的准确度，进而影响云量计算和云图分类的结果。本文的系统主要针对严重影响云检测准确度的两种噪声——星光噪声和日、月光噪声进行云图去噪，可见图 5(a) (b)。

2.2 云图分割

本文采用 Otsu 算法进行云区域检测，云区域标记为 1，天空背景标记为 0，得到二值图像，结果如图 5(c)所示。在此基础上将二值图像与原图相乘可分割出云区域，然后将分割图进行灰度归一，消除云图之间的明暗差异，最终云图分割结果如图 5(d)所示。

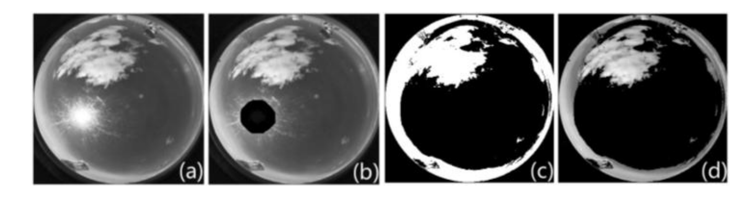


图 5 (a)原图（灰度图）；(b)去噪结果；(c)云区域检测；(d)云图分割

Fig.5 (a) Original image (grayscale image); (b) Denoising image; (c) Cloud area detection; (d) Cloud segmentation.

2.3 云量计算

系统在云区域分割的基础上，计算每张云图的 ASICDD 和 Scloud，作为下一步云图分类的特征。表 1 显示了数据集中部分云图的 ASICDD 和 Scloud，图 6 展示了数据集中云图的两个指标之间的关系，图中离差（deviation）计算公式如式 6 所示。可以看出，ASICDD 和 Scloud 在整体上呈正相关，经统计 76.44%的云图数据的离差在 0.1 以内，93.78%的云图的离差在 0.2 以内。但在云较少（Scloud<0.3）的情况下，二者并非严格的正相关，离差较大。出现 Scloud 较高但 ASICDD 较低的情况，这在表 1 列举出的样例中也有体现。这正是由于云分布密度的计算模型中加入了云点和中心点的距离影响：云面积大但距中心点远，云多分布在云图边缘，ASICDD 值低；反之，云面积小但云距中心点近，云多位于云图中央，则不利于天文观测，ASICDD 值高。ASICDD 和 Scloud 非正相关的情况多出现于 Clear 类和 Outer 类之中，可能会导致这两类之间的误判。

表 1 部分云图的云分布密度和云面积

Table 1 ASICDD and Scloud value of some ASI images.

	ASICDD	Scloud		ASICDD	Scloud
	0	0.0921		56.0984	0.5312
	2.0625	0.1543		59.5941	0.4799
	11.0089	0.0714		64.6411	0.5876
	22.1305	0.2822		75.2491	0.7054
	30.0241	0.3628		79.3264	0.7732

$$deviation = \left| Scloud - ASICDD/100 \right| \quad (6)$$

2.4 分类算法

这一部分本文采用了 SVM，KNN，决策树和随机森林 4 种传统算法进行分类测试。数据集中训练集包含 1056 张 ASI 图像，其中有 206 张 Clear，489 张 Covered，199 张 Inner 和 162 张 Outer。另外 455 张 ASI 图像用作测试集，包括 89 张 Clear，207 张 Covered，78 张 Inner 和 81 张 Outer。本文基于 Python 中的 Scikit-learn（Sklearn）对 ASI 图像进行分类，将计算的每张云图的 ASICDD 和 Scloud 值作为分类特征，训练集用于训练 4 个分类器，结合分层洗牌分割交叉验证训练超参数，使用准确率（Accuracy）作为模型评分。另采用文献[17]中的结合纹理特征与 SVM 的方法作对比，基于灰度共生矩阵（Gray-level Co-occurrence Matrix，GLCM）提取图像的灰度和纹理特征，结合 SVM 进行云图分类。

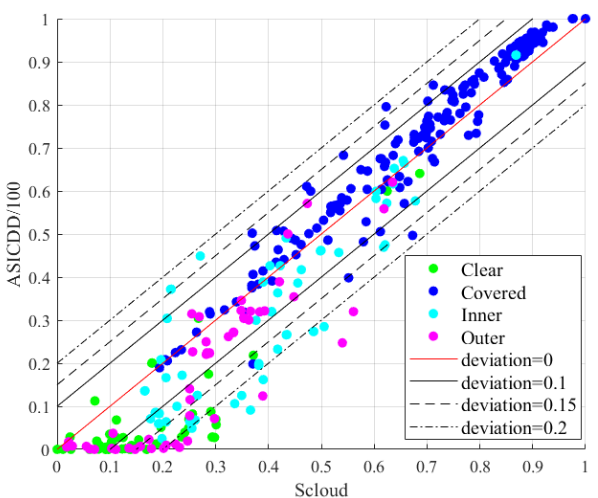


图 6 云分布密度和云面积之间的关系

Fig.6 The relationship of ASICDD and Scloud.

3 实验结果分析

实验部分是在为了较全面地评价本文提出的基于云分布密度和云面积的云图自动分类系统的分类效果，本文以分类准确率（Accuracy）、精准率（precision）、召回率（recall）和 F1-score 作为评判指标，在测试集 455 张 ASI 图像上进行自动分类，并与国家天文台提供的人工分类标签作对比，以人工分类结果为准，评估不同分类算法在各类云图上的分类性能。四个指标的定义如下：

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (7)$$

$$precision = TP/(TP + FP) \quad (8)$$

$$recall = TP/(TP + FN) \quad (9)$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

其中，TP 表示实为正并划分为正的样本数，TN 表示实为负并划分为负的样本数，FP 表示实为负却划分为正的样本数，FN 表示实为正却划分为负的样本数。测试集上的预测结果如表 2 所示。可以看出，本文使用的四种算法在各类云图的识别准确率上均达到了 90%以上，这表明本文提出的云分布密度和云面积可以作为可信的云量指标，在云图自动分类工作中展现了其优良性。总体准确率最高的是随机森林算法，Covered 和 Inner 达到了 100%准确识别。但 Outer 类准确率最高的是文献[17]中的方法，本文提出的方法很容易将 Outer 类误判为 Clear 类。基于 GLCM 特征提取的方法在 Outer 类的识别较准确，但对 Clear 类和 Inner 类的识别准确度较差，因此总体准确率不如本文提出的方法。

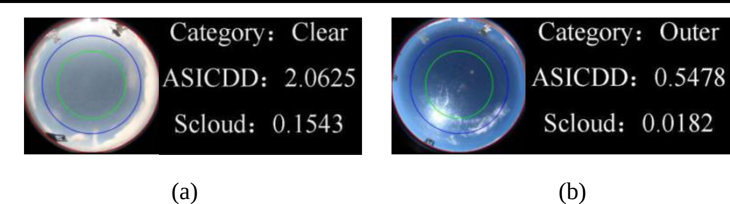
对于误判的云图，进行如图 7 所示的分析。图 7(a)是人工判定标签为 Clear，系统自动识别为 Clear 的云图；图 7(b)是人工判定标签为 Outer，系统自动识别为 Clear 的云图（误判云图）。可以看出，图 7(a)(b)在外圈以内均有云存在；图 7(a)中云较薄，分布较分散；图 7(b)中云较厚，分布较为集中。这就导致(a)图的云分布密度和云面积值均大于(b)图，因此导

致系统的误判。

表 2 自动分类准确率对比

Table 2 Comparison of accuracy of automatic classification results

Categories of cloud images		Clear	Covered	Inner	Outer	Total
Ours	SVM	0.9975	0.9807	0.9744	0.9012	0.9648
	KNN	0.9775	1	0.9872	0.9136	0.9780
	Decision Tree	0.9975	1	1	0.9072	0.9780
	Random Forest	**0.9975**	**1**	**1**	0.9506	**0.9868**
GLCM+SVM		0.9318	0.9903	0.9744	**0.9753**	0.9736



(a) (b)

图 7 云图误判分析

Fig.7 Misjudgment analysis of cloud images.

实验采用精准率（precision）、召回率（recall）和 F1-score 对分类结果进行评判，表 3 显示了四种分类器的分类结果评分，总体评分取样本的宏平均。可以看出，随机森林分类器的各项总评分为最高值，总精准率达到 98.41%，总召回率达到 98.2%，总 F1-score 达到 98.3%，在本文的自动分类系统中的分类性能最好。

表 3 使用精准率、召回率和 F1-score 评估分类结果

Table 3 Using precision, recall, and F1-score to evaluate classification results

Categories of cloud images		Clear	Covered	Inner	Outer	Total
precision	SVM	0.9158	0.9902	0.9500	0.9733	0.9573
	KNN	0.9355	0.9904	**1**	0.9737	0.9749
	Decision Tree	0.9457	0.9857	**1**	0.9733	0.9762
	Random Forest	**0.9667**	**0.9952**	**1**	**0.9747**	**0.9841**
	GLCM+SVM	0.9540	0.9951	0.9500	0.9634	0.9656
recall	SVM	**0.9775**	0.9807	0.9744	0.9012	0.9584
	KNN	**0.9775**	**1**	0.9872	0.9136	0.9696
	Decision Tree	**0.9775**	**1**	**1**	0.9012	0.9697
	Random Forest	**0.9775**	**1**	**1**	0.9506	**0.9820**
	GLCM+SVM	0.9326	0.9903	0.9744	**0.9753**	0.9681
F1-score	SVM	0.9457	0.9854	0.9620	0.9359	0.9573
	KNN	0.9560	0.9952	0.9935	0.9427	0.9719
	Decision Tree	0.9613	0.9928	**1**	0.9359	0.9725
	Random Forest	**0.9721**	**0.9976**	**1**	0.9625	**0.9830**
	GLCM+SVM	0.9432	0.9927	0.9620	**0.9693**	0.9668

4 结论

本文的创新点是提出了一个计算全天相机云图云量的直观可信的数值指标——云分布密度，并基于云量的数值特征建立云图自动分类系统，该系统实现了较高的识别准确率。本文测试了 SVM，KNN，决策树和随机森林 4 种传统算法；其中随机森林的准确率更高，各类云图的识别准确率达 95%以上。但是还存在云图误判的情况，需要进一步完善计算模型。此外，云图的光照条件容易影响云层厚度的判断，本文未能完全消除光照的影响。因此在接

下来的研究中需要考虑光照的影响，通过去噪消除光照影响，或是建立基于光照模型的云量计算模型。除此之外，还可以结合最新的机器学习算法完善系统，提高系统的识别准确率。

参考文献

[1] Dev S, Wen B, Lee Y H, et al. Ground-Based Image Analysis: A Tutorial on Machine-Learning Techniques and Applications[J]. IEEE Geoence & Remote Sensing Magazine, 2016, 4(2):79-93.

[2] Kotarba, Andrzej Z . Evaluation of ISCCP cloud amount with MODIS observations[J]. Atmospheric Research, 2015, 153:310-317.

[3] Long C N , Sabburg J M , Calbo J , et al. Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images[J]. Journal of Atmospheric and Oceanic Technology, 2006, 23(5):633-652.

[4] 杨俊，吕伟涛，马颖，等. 基于自适应阈值的地基云自动检测方法[J]. 应用气象学报，2009, 20(6):713-721.

Yang J, Lv W T, Ma Y, et al. An Automatic Ground-based Cloud Detection Method Based on Adaptive Threshold[J]. Journal of Applied Meteorology, 2009, 20(6):713-721.

[5] Liu S, Zhang L B, Zhang Z, et al. Automatic Cloud Detection for All-Sky Images Using Superpixel Segmentation[J]. Geoscience and Remote Sensing Letters, IEEE, 2015.

[6] Jelenak A , Heidinger A K . Validation of CLAVR-x cloud detection over ocean using AVHRR GAC sea surface temperature[J]. Proceedings of Spie the International Society for Optical Engineering, 2005, 5658:292-298..

[7] Gang, M. A., Zongyi, F., Fengying, Z.: An error analysis on the simulated data of TOVS HIRS by RTTOV5 model[J]. Journal of Meteorological Research, 2002,16(2):226-241.

[8] Cheng H Y , Lin C L . Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques[J]. Atmospheric Measurement Techniques, 2017, 10(1):1-20.

[9] Blazek M , Pata P . Colour transformations and K-means segmentation for automatic cloud detection[J]. Meteorologische Ztschrift, 2015, 24(5):503-509.

[10] Taravat A , Del Frate F , Cornaro C , et al. Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images[J]. IEEE Geoence & Remote Sensing Letters, 2014, 12(3):666-670.

[11] Dev S, Member, IEEE, et al. CloudSegNet: A Deep Network for Nychthemeron Cloud Image Segmentation[J]. IEEE Geoscience & Remote Sensing Letters, 2019, PP(99):1-5.

[12] Ye L, Cao Z, Xiao Y, et al. Supervised Fine-Grained Cloud Detection and Recognition in Whole-Sky Images[J]. IEEE Transactions on Geoence and Remote Sensing, 2019, PP(99):1-14.

[13] Dev S, Savoy F M, Lee Y H , et al. High-Dynamic-Range Imaging for Cloud Segmentation[J]. Atmospheric Measurement Techniques, 2018, 11(4).

[14] 张家强，李潇雁，李丽圆,等．基于深度残差全卷积网络的 Landsat 8 遥感影像云检测方法[J]．激光与光电子学进展, 2020, 57(10):364-371.

Zhang J Q, Li X Y, Li L Y, et al. L andsat8 remote sensing image based on deep residual fully convolutional network[J]. Laster & Optoelectronics Progress, 2020,57(10):364-371.

[15] 郭玥，于希明，王少军,等．遥感图像云检测的多尺度融合分割网络方法[J]．仪器仪表学

报, 2019,40(6):31-38.

Guo Y, Yu X M, Wang S J, et al. Cloud detection in remote sensing images with multilevel scale fused network[J]. Chinese Journal of Scientific Instrument, 2019,40(6):31-38.

[16] Shi C J, Zhou Y T, Qiu B, et al. Diurnal and nocturnal cloud segmentation of all-sky imager (ASI) images using enhancement fully convolutional networks[J]. Atmospheric Measurement Techniques, 2019, 12(9):4713-4724.

[17] 贾杰,邵丽群.遥感领域中气象卫星云图自动分类研究[J].无线互联科技, 2016(01):120-123.

Jia J, Shao L Q. The research of automatic classification of meteorological satellite imagery in remote sensing field[J]. Wireless Internet Technology, 2016(01):120-123.

[18] 魏诗雅,邱波,曹子皇,等.一种全天相机云图的自动处理方法[J].天文研究与技术, 2019,16(01):85-92.

Wei S Y, Qiu B, Cao Z H, et al. An Automatic Processing Method for All-sky Images[J]. Astronomical Research & Technology, 2019,16(01):85-92.

[19] 高振斌,李梦慈,邱波,等.全天相机云图去噪算法的研究[J].天文研究与技术, 2020,17(01):76-84.

Gao Z B, Li M C, Qiu B, et al. Research on denoising algorithm for camera image of all-day camera[J]. Astronomical Research & Technology, 2019,16(01):85-92.

A cloud cover measurement index based on all-sky imager images

Zhang Yuxin[1], Qiu Bo[1], Shi Chaojun[1], Li Mengci[1], Xiang Guanjie[1]
(1.Hebei University of Technology，Tianjin，300401)

Abstract: All-sky imager (ASI) images are an important means for the astronomy community to monitor cloud cover at present, but there is no definite cloud cover calculation index. Therefore, this paper proposes a new quantitative index for cloud amount measurement--Cloud Distribution Density of ASI images (ASICDD), and establishes an automatic classification system for ASI images based on this index. Firstly, all the images in the dataset from the Key Laboratory of Optical Astronomy at the National Astronomical Observatories of Chinese Academy of Sciences (CAS) are denoised and the cloud area is segmented from the sky by OTSU algorithm. Secondly, the cloud amount for ASI image with the background remove is calculated by ASICDD proposed in this paper. Finally, we use four traditional classifiers (SVM, KNN, Decision Tree and Random Forest) to classify ASI images automatically according to the calculated value and evaluate the performance of each classifier. The results show that ASICDD can be used as a numerical index for judging the cloud cover of ASI images; the automatic ASI images classification system based on ASICDD achieves a high recognition accuracy rate. Meanwhile Random Forest has the best classification effect--the recognition accuracy rate of various cloud images has reached more than 95%.

Keywords: ASICDD; all-sky images; TMT; cloud cover measurement